

Case ID:M14-184P

Published: 12/22/2015

## Inventors

**Werner Dahm**

## Contact

Shen Yan  
shen.yan@skysonginnovations.  
com

## Improved Data-Centric Classification Method Including Application to Predictive Risk Scoring

Big data analytics uncover hidden patterns, unknown correlations, and other practical information that can be applied to make more accurate predictions and decisions. Current techniques rely on statistics and decision-tree algorithms in order to “mine” useful information from massive data sets that are largely dominated by irrelevant data points. These irrelevant data points provide little to no useful information about the relationship between data sets, creating background “noise” that weakens any relevant correlations and increases the error within the statistical models used by analytical software. Weaker correlations among the relevant data means that significant numerical relationships go undetected, so riskier clients are more likely to be approved and fraudulent or threatening behavior is less likely to be identified.

Researchers at ASU have developed a method that dramatically improves comparisons between a given data set and two or more other data sets, even when the data sets differ in size or are grouped in different locations relative to one another. The method works by partitioning each data set over a common domain (resulting in equal dimensions necessary for subtraction), subtracting out related data points, and comparing the remaining differences. For example, a data set representing known normal behavior would be subtracted from a data set representing known malicious behavior and from the data set in question. The two resulting data sets exclude the unnecessary data that contributes to background noise while retaining their useful information. This method does not interfere with standard procedures for dimensionality reduction and hypotheses can be still tested using ordinary statistical techniques. This method facilitates far more accurate analytics with minimal modeling error, leading to fewer operational risks and earlier fraud or threat detection.

### Potential Applications

- Bank Security
- Forecasting
- Machine Learning
- Risk Assessment
- Underwriting

### Benefits and Advantages

- Accurate
  - Distinctly expresses relevant data points for more sensitive comparisons.
  - Lower levels of background noise reduces error in statistical models.
- Innovative – Risks are better averted and suspicious behavior is caught earlier.

- Retrofit – Can be applied to and used in conjunction with existing methods.
- Versatile – Works even when data sets differ in size and relative location.

For more information about the inventor(s) and their research, please see

[Dr. Werner J.A. Dahm's directory webpage](#)